Video analytics for retail

A.W. Senior, L. Brown, A. Hampapur, C.-F.Shu, Y. Zhai, R.S. Feris, Y.-L. Tian, S.Borger, C.Carlson
aws @ us.ibm.com
IBM T. J. Watson Research Center, PO Box 704,
Yorktown Heights, NY 10598, USA.

Abstract

We describe a set of tools for retail analytics based on a combination of video understanding and transaction-log. Tools are provided for loss prevention (returns fraud and cashier fraud), store operations (customer counting) and merchandising (display effectiveness). Results are presented on returns fraud and customer counting.

1. Introduction

Closed Circuit Television (CCTV) has long been used within shops for the detection of shoplifting. CCTV systems have proved to have a variety of uses to justify investment— as deterrent, record for insurance claims, public safety, stock tracking and employee fraud detection— but they are still labour intensive and it is difficult to extract useful information from them. Historically cameras have been steered by loss prevention staff to track suspected shoplifters, achieving high-resolution covert observation. Such streams are recorded to provide evidence for convictions. In addition, cameras not being actively controlled may be recorded to provide a record of activities in important areas, such as entrances and high-value item displays. The advent of digital video recorders has dramatically improved the access to this recorded video enabling faster investigation of past events by direct access based on recording time, when the time of an event is known.

The development of intelligent video processing algorithms is bringing many new applications for video within stores, both in the traditional domain of loss prevention, and in store operations and merchandising. This paper describes a set of tools for retail analysis based on video tracking algorithms (Section 2.1). The system provides functions to assist with everyday loss-prevention video surveillance and for the investigation of cashier fraud using point of sale transaction logs (TLOGs) to index into video (Section 3.2.1) with a specific tool for the investigation of returns fraud (Section 3.2.2). It also provides tools for the estimation of customer traffic into and out of the store, within individual departments (Section 3.3.2) and for measuring the effectiveness of a particular display, based on the traffic and customer behaviour (Section 3.3.3).

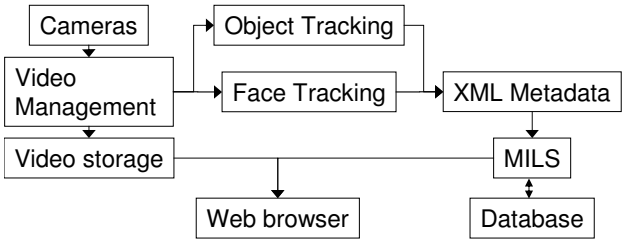

Figure 1: Schematic showing the main system components.

1.1. Related work

Several previous works, and three others in this volume have addressed video processing in retail environments, indeed the PETS 2002 [1] workshop was based around video recorded in a shopping mall, with tasks of counting people passing and standing in front of a shop window. Two papers have described the problem of determining when groups of people are together (*e.g.* a family), constituting a single "shopping group" likely to make a joint purchase, and thus more indicative of sales traffic than is the number of people observed. Haritaoğlu *et al.* [4] described a system for counting shopping groups waiting in checkout lanes and Leykin *et al.* [6] have used swarming algorithms to group customers tracked throughout a store into shopping groups.

Several companies offer video-based people tracking solutions for retail environments, from people counting at entrances and tracking throughout a store (e.g. Brickstream, ShopperTrak) These solutions tend to be designed around top-down camera views useful only for the specialized vision system, sometimes requiring stereo cameras, as in [3]. Wolfe *et al.* [13] made a system that used low-cost, low-resolution IR sensors for through-store tracking. Haritaoğlu and Flickner [5] have also examined the use of computer vision systems to measure shoppers' attention to a product promotion or billboard, counting the number of people and the time spent observing the display. Active lighting is used to extract eye contact, and an SVM is used to determine peoples' gender. Liu *et al.* [7] have used active appearance models to determine head gaze for the same purpose.

Venetianer *et al.* [12] have explored the coupling of TLOG records from point of sales terminals with video analytics for the detection of suspicious conditions. In a previous paper [11] we have coupled TLOG with video event detection to detect returns fraud in a department store.

2. Surveillance system

The retail applications are based on our video surveillance system (see schematic in Figure 1) which is a distributed computing system to extract and filter meaningful content from the deluge of video from multiple streams, and make available meaningful alerts, reports and visualisations based on this data. A video management subsystem handles the routing and storage of video, which is generally digitised and compressed at source and distributed over TCP/IP. Video processing servers run the Smart Surveillance Engine (SSE) which incorporates a variety of video analytics configured independently for each video feed. The SSE generates metadata which is transmitted as XML via TCP/IP to a Server running the Middleware for Large scale Surveillance (MILS). This backend ingests the content through a web services interface and stores the index and content in a relational database. MILS also provides a web services API to deliver content to application clients, which may be standalone programs but in this instance are AJAX applications running in a standard web browser.

2.1. Video analytics

The main video analytics algorithms offered by the systemare generic object tracking and face tracking. Both trackers send object appearance, trajectory and keyframe information to the database.

The face tracking algorithm [8] uses a cascade of feature detectors to detect faces (either frontal or profile) and a correlation tracker to maintain hypotheses when not detected by the detector. Adaboost learning is used to obtain the cascade of classifiers, using a feature pool based not only on the traditional rectangle features, but also on Gabor wavelets optimized to match the local geometric structure of the training samples. Real-time processing (20-25Hz) is achieved by interleaving multiple view-based detectors (frontal and profile) in the temporal domain.

When faces are not clearly visible, customers are tracked by the *ColourField* tracker, run independently on each camera. First background subtraction produces a foreground mask indicating moving objects that are not explained by the background model. We use a fast multiple-Gaussian algorithm [2] that provides robustness to changes of lighting and shadows. The detected foreground regions are tracked with a probabilistic appearance model tracking algorithm[10]. This models the shape and appearance of objects and allows pixel-wise resolution of occlusions of multiple objects, with continuous identity maintenance of objects during visual occlusions.

3. Retail analytics users

The guiding design principle in the development of the retail system was the understanding that users of a video analytics system will have a variety of needs which depend on their role. Currently CCTV systems are used mainly by Loss Prevention staff whose role is to prevent shrinkage (see Section 3.1). The most frequent users spend much of their time observing current shoppers and looking for suspicious behavior. When suspicious behaviour is detected the employee will continue observation of the suspicious person(s) until their suspicions are allayed, or the person(s) leave the store. If a theft has been observed then the suspects are apprehended by employees, who in a small store, may be the same people who made the video observations. Direct observation may also be used when internal fraud is suspected with cameras being used to observe employees at points of sales or store entrances. Section 3.1.1 describes how video analytics can be used to assist loss prevention staff in this "live monitoring" role.

Another activity of loss prevention staff is the "off-line" investigation of shrinkage — trying to detect the causes, perpetrators and volumes of past shrinkage. This task, involves the assimilation of data from a number of sources, such as TLOGs, shipping and staffing records, and increasingly from video recordings. This "Loss Prevention Manager" role is supported by a number of functions which are described in section 3.2.

New uses of video are being developed besides loss prevention, particularly in store operations and merchandising. Store operations encompasses a wide variety of activities, many of which can be aided by video analytics, from planning store layouts based on customer path statistics to staff planning based on historical and instantaneous customer counts, at store entrances, departments and checkout queues. Merchandising activites can also be planned based on similar analytics- choosing the location of a display based on customer paths, as well as measuring the effectiveness of a display based on customer counts coupled with sales figures. Section 3.3 describes video analytics for store operations and merchandising which are implemented in the "Store manager" role.

3.1. Loss Prevention

"Shrinkage" is a catch-all term to describe a shortfall in the accounts of retail stores, unnecessary loss which businesses are keen to reduce. Stores in developed countries may have a shrinkage of 1–2% of sales [9], measured by subtracting stock levels and sales from deliveries. The detailed causes of shrinkage are usually unknown. In many stores there is a "Loss Prevention" department whose role is to reduce shrinkage. Shrinkage may include:

- Clerical error (miscounting stock, accounting errors)
- Misplaced, undelivered or "lost" stock
- Shoplifting
- Employee theft
- Theft by supplier
- Returns fraud
- Tag switching (putting a lower price tag on an item)
- Sweethearting (employee-customer collusion)

3.1.1 Live monitoring

Most time of loss prevention staff is currently spent monitoring people in the store, mostly through CCTV, for the prevention and detection of shoplifting and fraud by customers and employees. Operators observe customers and employees to look for suspicious "indicator behaviours" and then attention is focussed on the more suspicious cases.

An automatic video analytics system can help in this live monitoring task in a number of ways, foremost by monitoring many channels of video simultaneously for indicator behaviours. Video analytics are not yet sufficiently refined to pick up all the indicator behaviours that trained LP staff may detect, including concealing merchandise, tag switching or looking around suspiciously. However, video analytics can detect events such as the following: when people enter low-traffic areas that may be used for tag-switching and merchandise concealment; any motion in front of the safe; any person close to a fire door; out-of-hours activity in restricted areas, such as loading docks. Figure 2 shows automatic alerts generated when a person enters or leaves a changing room, with keyframes providing a side-by-side comparison of merchandise and bags being carried in and out that can be compared instantly without constantly watching the video. Clicking on a keyframe plays the original video for more detail. Such automated alerts are displayed automatically in the user interface, and can be used to drive the workflow of LP staff by directing their attention at potential indicator behaviour from many different cameras when not engaged in a higher priority investigation. A video analytics system can thus be integrated with Electronic Article Surveillance (EAS) tags and other sources of intelligence, into a workflow management system that presents and prioritizes shrinkage risks for investigation by loss prevention staff.

Figure 2: Images of a customer entering and leaving a changing room, detected by a motion detection region.

3.1.2. Moving cameras

One important feature rarely considered in automated surveillance systems is the fact that existing infrastructure depends on pan-tilt-zoom cameras, for maximum coverage with a small number of cameras, but automatic algorithms mostly work only on static visual feeds. Requiring static cameras for analytics algorithms entails the installation of additional cameras which may be less useful to LP staff than a PTZ camera in the same location.

To enable our system to exploit existing infrastructure as much as possible, it is designed to operate on piecewise static cameras — cameras which are stationary except when they are being actively used by LP staff to track a person.

In a background-subtraction system, camera movement results in much of the image being detected as foreground and the consequent failure of tracking. To avoid these problems, incoming images are first processed by a camera motion detection algorithm which estimates the camera movement by point tracking with a RANSAC motion estimation. Small vibrations are compenstated by shifting the image resulting in a stabilized output. Larger movements are detected and trigger suspension of normal background subtraction and tracking, as well as sending an alert. While tracking is suspended, estimation of the type and extent of the motion continues. When motion ceases, a further alert is stored in the database, labelled with the attributes of the camera motion (e.g. zoom in $\times 2$ and pan left $40°$). Detection and tracking are resumed in the new location.

False positive detections due to camera motion are thus avoided, and the database contains a set of indices which are useful for LP staff to search for incidents, since they can find occasions when they were actively steering a camera. Finally, for cameras that are intended to provide continuous coverage of a specific area, we enable the camera's "home on timeout" feature so that when not being actively used by LP staff, the camera returns automatically to the default position and resumes its main monitoring function.

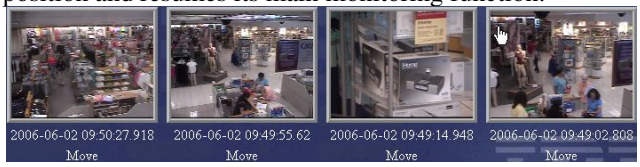

Figure 3: Database alerts generated by moving a camera

3.2. Loss prevention manager

As mentioned in section 3, loss prevention staff may pursue off-line, post hoc investigation of shrinkage. Such investigations vary from finding the events leading up to a shoplifter running out of the store, to a detailed investigation of whether store policies were followed in the issuing of coupons. Often these investigations are driven by knowing the time of an incident, or looking it up based on a source such as the TLOG. Hitherto the LP manager may have had to pull a tape and fast-forward to the desired time or manually enter a time on a digital video recorder.

3.2.1. Transaction Log Integraion

Our system integrates a TLOG feed from either the store points of sales or a central database. Through an integration with IBM's Store Integration Framework, TLOG events are

transferred to the MILS server immediately after they occur. Alternatively, a periodic feed of batched events from the store or head office information system is delivered to an ingestion server that converts the TLOG events into XML and transmits them to the MILS server.

TLOG data can be browsed and searched in a number of ways, displaying transaction statistics over hours, days or months and drilling down or searching for individual transactions using time or any other fields of interest, including item value, name or SKU; cashier name or ID; register; and transaction type (dependent on the store, but typically including sale, void, tender, price inquiry, return, log on/off, supervisor authorisation etc.). Voids and returns in particular can be indicators for frauds.

Figure 4 shows the TLOG search interface showing individual returns transactions for a particular register. Each transaction is hotlinked in a number of ways. Individual fields are linked to search for transactions sharing common attributes, and the "Register video" and "Customer video" columns contain links to video from two cameras that may be available — one of the register itself for observing cash drawer and cashier activity, and one (potentially the same) of the customer which will also show items being bought.

3.2.2. Returns Fraud

Returns fraud can take one of a number of forms. One of these is the return of items that are not eligible for return (broken, out of policy window) but a more serious problem is that of returning items that were never bought, either returning an item that was just picked up in the store without a receipt (in stores that have liberal returns policies), or using a receipt from a previously purchased (and kept) item to return a new item just picked up in store.

A number of possible solutions present themselves before considering video: a stricter returns policy, the requirement for a receipt, placement of customer service at the front of the store, unique serial numbers scanned at purchase (rather than product-type codes) or even RFID tracking of items. All of these methods have drawbacks, principally cost and operational complexity, but also fears of impact on customer satisfaction. Thus we offer an unobtrusive solution that exploits existing video infrastructure integrated with other video analysis functions.

Our approach, described in more detail in a previous paper [11], allows loss prevention staff to quickly determine whether a person returning an item entered the store carrying that item. The system works by detecting and tracking customers at entrances and customer service desks and associating the two events. Such a solution only requires cameras at the store entrances and returns counters, so is simpler than an approach where the customer is tracked throughout the store (requring many cameras and very reliable camera hand-off algorithms) and must be continuously monitored to determine whether items are picked up.

Two cameras at customer service record activity there, including the appearance of customers returning items. A separate set of cameras points at the doors and capture all activity of people entering and leaving the store. Figure 5 shows the fields of view of two such cameras.

Figure 5: Views from the customer service desk (left) and a door (right). The "region of uninterest" in which detections are ignored to reduce false positives is outlined in blue. Alert tripwires (*enter* and *leave*) are drawn on the door view.

Our approach to returns fraud is to segment automatically events in each of these cameras, to filter them and then provide a user interface which allows the association of each returns event with the door entrance event showing when the person came into the store. At the customer service desk, the face tracking algorithm (Section 2.1) tracks customers' faces, generating one event per customer. Customers at the doors are tracked with the ColourField tracker. Detecting entrance events from the store doors is a challenging task, because of lighting, geometry and the presence of distracting motion (particularly of the doors). Alerts are set up to allow particular events to be detected among the movements observed by the door cameras. Directional tripwires are drawn in front of each door, with tracks crossing the tripwires are flagged as "exit" or "enter" events. In the returns fraud application, only those people entering the store are displayed. Other events are stored for searching and visualization through other interfaces (Section 3.3.1).

The returns fraud interface provides intuitive selection and browsing of the events, summarized by presentation of keyframes (at both scales), timestamps and original video clips (from DVR or media server). Search typically begins by selecting a return event from the TLOG. In response to this, the interface displays people found at the customer service counter near that time. Selecting one of these then displays people entering the store shortly before the selected event. The user can then browse through the entrance events to find a match, using full-frame and zoomed in keyframes as well as original video to make the comparison and, when a match is found, to determine if fraud has taken place.

The fundamental indexing attribute of the database is time. All devices are synchronized and events are timestamped. Temporal constraints from real world conditions are exploited to limit the events displayed. Empirically, we find that 70% of people take between 1 and 3 minutes to walk from the entrance to customer service and few peo-

Figure 4: Interface for browsing and searching TLOG showing a table of returns transactions.

ple take longer than 20 minutes. Thus in the vast majority of cases, only a few entrance events need be searched for a match. An archive button allows the user to save the matched events for rapid future access, and preserves these events from automatic data expiration.

3.3. Store Manager

As indicated in Section 3, the store manager account provides access to both store operations and merchandising functions of the system. The store manager is provided with the same interface for searching and browsing TLOGs, with the ability to graph trends over days or months. In addition, the store manager can access interface pages to investigate traffic flow in the store and display effectiveness.

3.3.1. Customer counting

The number of customers entering a store is one of the most important statistics of interest to a store manager that cannot be obtained from the TLOG. Several specialized methods for counting customers are available (beambreakers, pressure pads, or staff with click counters) but all have drawbacks in accuracy or expense. In this system customer counting is carried using video analytics to count people passing through known entry or exit points. Stores generally have cameras directed at the doors to observe all entrance and exit traffic, and we use such existing cameras in this system although their oblique angle (suitable for identifying customers in the returns fraud application, but subject to significant occlusions) makes counting customers less accurate than would be possible with top-down cameras. The alerts described in Section 3.2.2 trigger each time a person exits or leaves the store, and the resulting counts can be graphed over time, as shown in Figure 6, with the ability to drill down to each customer's keyframe and video. The entrance and exit counts can be used in conjunction to estimate the number of customers in the store at any time and the average time spent in the store.

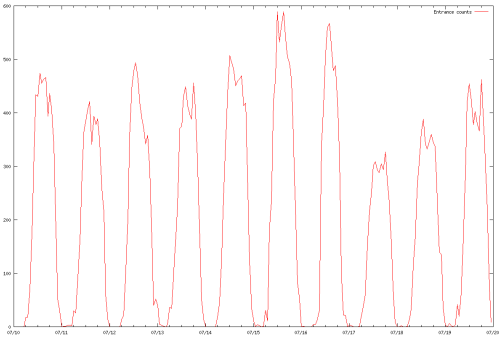

Figure 6: Hourly (entering) customer counts over 10 days.

3.3.2. Traffic flow

Almost as important as counting the number of people in the store is knowing where people go within the store. To this end the interface includes a department or camera level traffic analysis tool. This graphs the traffic in a particular camera over time, and can also visualize results both by drawing the tracks themselves or a "heat map" (Figure 7) showing areas with greatest activity.

Figure 7: A colour density plot of customer activity.

3.3.3. Display effectiveness

Display effectiveness is evaluated in the interface by calculating statistics on where customers spend their time in a

view of a display. The interface allows the user to choose active regions, such as the area in front of a display, and observe how many customers entered the region in a period of time, how many stopped there and how long these customers spent. All the trajectories of customers are shown and allow the user to "drill down" to the original video to observe the behavior of the selected customers.

Figure 8: The display effectiveness view, showing customer trajectories and statistics for interaction with the display.

3.3.4. Conversion rate

The store wide counts can also be used to calculate average sales per customer or "conversion rate". The number and value of sales in each hour can be calculated from the TLOG. Dividing by the number of customers leaving the store (leaving times will correlate more closely with sales than entrance times), number and value of sales per person entering can be calculated, and graphed in the interface. Local people counts can also give estimates of conversion rates per store or per display.

4. Results

Systems based on the functionality described here have been deployed at a number of stores in the US and South America and have been in operation for over a year.

Much of the use of the system is qualitative, allowing the observation of individual instances and trends. Analysis of the system performance has been focused on returns fraud. Several days of video data were recorded from 6 cameras (4 doors, 2 returns counters) and processed by the system. Several users carried out the full returns fraud people search task, with or without TLOG information, and compared ingested data to hand-marked ground truth of entrance events.

The following measures were evaluated:

- Overall match proportion: Proportion of customers at customer service found at the entrance: 85%
- Overall match time: Average time to find a match using the interface: 86s
- Customer detection: Proportion of customers at customer service detected and displayed in interface: 85%
- Entrance detection: Proportion of people entering the store visible in the event keyframes: 95%

The system deployed in the store uses dual 3.6GHz Pentium servers for video analytics, video management, and MILS. The ColourField tracking algorithm runs at thirty-frames per second. Background subtraction takes between 5.5 and 8.5ms per frame and tracking taking between 2 and 4 ms when there is foreground to be tracked.

5. Conclusions

This paper has described a practical, versatile tool for video applications in a retail store. The system uses detection, tracking and indexing capabilities of a generic video surveillance system to allow a comprehensive set of reporting and loss prevention applications, all made available remotely through a standard web browser. Investigations can be driven by point of sale transation logs or video events, with events always being linked back to original video. A video-based system demonstrates advantages over dedicated systems (for people counting, say) in that the infrastructure serves multiple functions and provides rich data where needed in addition to statistics and discrete events.

References

[1] *IEEE Workshop on Performance and Evaluation of Tracking and Surveillance Systems*, 2002.

[2] J. Connell, A.W. Senior, A. Hampapur, Y.-L. Tian, L. Brown, and S. Pankanti. Detection and tracking in the IBM PeopleVision system. In *IEEE ICME*, June 2004.

[3] I. Haritaoğlu, D. Beymer, and M. Flickner. Ghost3D: Detecting body posture and parts using stereo. In *Workshop on Motion and Video Computing*, pages 175–80. IEEE, 2002.

[4] I. Haritaoğlu and M. Flickner. Detection and tracking of shopping groups in stores. In *Conference on Computer Vision and Pattern Recognition*, pages 431–438, 2001.

[5] I. Haritaoğlu and M. Flickner. Attentive billboards: Towards to video based customer behavior understanding. In *Proc. IEEE Workshop on Applications of Computer Vision*, 2002.

[6] A. Leykin and M. Tuceryan. Detecting shopper groups in video sequences. In *This volume*.

[7] X. Liu, N. Krahnstoever, T. Yu, and P. Tu. What are customers looking at? In *This volume*, 2007.

[8] Y. Tian R. Feris and A. Hampapur. Capturing people in surveillance video. In *IEEE International Workshop on Visual Surveillance*, Minneapolis, MN, 2007.

[9] Centre For Retail Research. The European retail theft barometer. Technical report, www.retailresearch.org, 2005.

[10] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance models for occlusion handling. In *International Workshop on Performance Evaluation of Tracking and Surveillance*, 2001.

[11] A.W. Senior, L. Brown, C.-F. Shu, Y.-L. Tian, M. Lu, Y. Zhai, and A. Hampapur. Visual person searches for retail loss detection. In *Intl. Conf. on Vision Systems*, 2007.

[12] P. Venetianer, Z. Zhang, A. Scanlon, Y. Hu, and A. Lipton. Video verification of point of sale transactions. In *This vol.*

[13] R. H. Wolfe, P. C. Hobbs, and S. Pankanti. Footprints: An IR approach to human detection and tracking. In *Proc. SPIE*, volume 4554, pages 42–51, September 2001.